

PATENT
0512-1176

IN THE U.S. PATENT AND TRADEMARK OFFICE

In re application of: Samir NEFTI et al.
Conf.:
Appl. No.:
Group:
Filed: August 12, 2003
Examiner:
Title: METHOD OF AUTOMATIC PROCESSING OF A
SPEECH SIGNAL

CLAIM TO PRIORITY

Assistant Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

August 12, 2003

Sir:

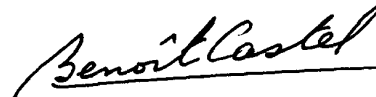
Applicant(s) herewith claim(s) the benefit of the
priority filing date of the following application(s) for the
above-entitled U.S. application under the provisions of 35
U.S.C. § 119 and 37 C.F.R. § 1.55:

<u>Country</u>	<u>Application No.</u>	<u>Filed</u>
FRANCE	02 13417	October 25, 2002

Certified copy(ies) of the above-noted application(s)
is(are) attached hereto.

Respectfully submitted,

YOUNG & THOMPSON



Benoit Castel, Reg. No. 35,041

745 South 23rd Street
Arlington, VA 22202
Telephone (703) 521-2297

BC/ia

Attachment(s): 1 Certified Copy(ies)



BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 11 JUIN 2003

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS cedex 08
Téléphone : 33 (0)1 53 04 53 04
Télécopie : 33 (0)1 53 04 45 23
www.inpi.fr



**BREVET D'INVENTION
CERTIFICAT D'UTILITÉ**

Code de la propriété intellectuelle - Livre VI


N° 11354*02
26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08

Téléphone : 33 (1) 53 04 53 04 Télécopie : 33 (1) 42 94 86 54

REQUÊTE EN DÉLIVRANCE**page 1/2****BR1**

Cet imprimé est à remplir lisiblement à l'encre noire

DB 5-20 W / 010801

REMISE DES PIÈCES DATE 25 OCT 2002 LIEU 75 INPI PARIS N° D'ENREGISTREMENT 0213417 NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE 25 OCT. 2002 PAR L'INPI		1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE CABINET LAVOIX 2, Place d'Estienne d'Orves 75441 PARIS CEDEX 09	
Vos références pour ce dossier BFF 02/0363 <i>(facultatif)</i>			
Confirmation d'un dépôt par télécopie <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
2 NATURE DE LA DEMANDE Demande de brevet Demande de certificat d'utilité Demande divisionnaire <i>Demande de brevet initiale</i> <i>ou demande de certificat d'utilité initiale</i> Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i>		Cochez l'une des 4 cases suivantes <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> N° _____ Date _____ N° _____ Date _____ N° _____ Date _____	
3 TITRE DE L'INVENTION (200 caractères ou espaces maximum) Procédé de traitement automatique d'un signal de parole.			
4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation _____ N° _____ Date _____ Pays ou organisation _____ N° _____ Date _____ Pays ou organisation _____ N° _____ Date _____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
5 DEMANDEUR (Cochez l'une des 2 cases) Nom ou dénomination sociale Prénoms Forme juridique N° SIREN Code APE-NAF Domicile ou siège Rue Code postal et ville Pays Nationalité N° de téléphone <i>(facultatif)</i> Adresse électronique <i>(facultatif)</i>		<input checked="" type="checkbox"/> Personne morale <input type="checkbox"/> Personne physique FRANCE TELECOM Société Anonyme _____ _____ 6, Place d'Alleray 75015 PARIS FRANCE Française N° de télécopie <i>(facultatif)</i> _____ <input type="checkbox"/> S'il y a plus d'un demandeur, c. chez la cas et utilisez l'imprimé «Suite»	

Remplir impérativement la 2^{ème} page

REMISE DES PIÈCES DATE LIEU N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI		Réservé à l'INPI 25 OCT 2002 75 INPI PARIS 0213417	DB 540 W / 010801
Vos références pour ce dossier : <i>(facultatif)</i>		BFF 02/0363	
6 MANDATAIRE <i>(s'il y a lieu)</i> Nom Prénom Cabinet ou Société N° de pouvoir permanent et/ou de lien contractuel Adresse Rue Code postal et ville Pays N° de téléphone <i>(facultatif)</i> N° de télécopie <i>(facultatif)</i> Adresse électronique <i>(facultatif)</i>		CABINET LAVOIX 2 Place d'Estienne d'Orves 75441 PARIS CEDEX 09 FRANCE 01 53 20 14 20 01 48 74 54 56 brevets@cabinet-lavoix.com	
7 INVENTEUR (S) Les demandeurs et les inventeurs sont les mêmes personnes		Les inventeurs sont nécessairement des personnes physiques <input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non : Dans ce cas remplir le formulaire de Désignation d'inventeur(s)	
8 RAPPORT DE RECHERCHE Établissement immédiat ou établissement différé Paiement échelonné de la redevance <i>(en deux versements)</i>		Uniquement pour une demande de brevet (y compris division et transformation) <input checked="" type="checkbox"/> Établissement immédiat <input type="checkbox"/> Établissement différé Uniquement pour les personnes physiques effectuant elles-mêmes leur propre dépôt <input type="checkbox"/> Oui <input type="checkbox"/> Non	
9 RÉDUCTION DU TAUX DES REDEVANCES Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention <i>(joindre un avis de non-imposition)</i> <input type="checkbox"/> Obtenue antérieurement à ce dépôt pour cette invention <i>(joindre une copie de la décision d'admission à l'assistance gratuite ou indiquer sa référence)</i> : AG <input type="text"/>	
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire)		VISA DE LA PRÉFECTURE OU DE L'INPI M. ROCHET	

La présente invention concerne un procédé de traitement automatique d'un signal de parole comportant une étape d'alignement entre une modélisation d'un texte et un signal de parole correspondant à l'élocution de ce texte.

De tels procédés sont utilisés, par exemple, dans le cadre de la synthèse de parole ou encore de la détermination de bases de données d'apprentissage pour des systèmes de reconnaissance vocale.

En référence à la figure 1, on va décrire un procédé de traitement automatique de l'état de l'art dans le cadre de la synthèse vocale.

Ce procédé comporte une étape 2 automatique de détermination d'une séquence de modèles probabilistes représentative d'un texte donné.

De manière classique, les modèles probabilistes utilisés sont un nombre fini de modèles dits « modèles de Markov cachés » ou HMM (Hidden Markov Model) qui décrivent la probabilité de réalisation acoustique d'unités symboliques de nature phonologique.

Parallèlement à l'étape 2, le procédé comporte une étape 4 de détermination d'une séquence de trames de données numériques correspondant à l'élocution du même texte donné, ou trames acoustiques.

Le procédé comporte ensuite une étape 6 d'alignement entre la séquence de trames acoustiques et la séquence de modèles.

Ainsi, à chaque unité symbolique d'ordre phonologique représentée par un ou plusieurs modèles, est associée une sous-séquence de trames acoustiques dite « segment acoustique ».

Par exemple, ces associations entre une unité symbolique et un segment acoustique sont mémorisées individuellement afin de permettre ultérieurement la synthèse de parole en générant une séquence de trames acoustiques correspondant à un autre texte que le texte donné précité.

Cependant, des variations peuvent apparaître lors de l'étape 6 d'alignement résultant notamment de différences entre le signal de parole réellement prononcé et la séquence de modèles correspondant à une prononciation théorique.

En effet, l'étape 2, de détermination d'une séquence de modèles, associe à un texte donné une unique séquence de modèle.

Cependant, l'élocution de ce texte peut donner lieu à différents signaux de parole du fait de l'influence du locuteur. Notamment, des unités phoné-

tiques, ou phonèmes peuvent être associés entre eux comme dans le cas de liaisons, ou encore d'autres phonèmes peuvent être supprimés ou rallongés.

De telles variations peuvent entraîner l'association d'un modèle avec un segment acoustique erroné et/ou décalé, introduisant ainsi une erreur
5 d'alignement dans les segments acoustiques suivants.

Il résulte de ces variations, la nécessité d'introduire pour chaque association entre un segment acoustique et un ou plusieurs modèles, un indice de confiance, lors d'une étape 8, qui permet d'attribuer un score probabiliste à chaque association.

10 Cependant, dans les procédés de l'état de l'art, ces indices de confiance calculés pour chaque modèle ne sont pas d'une grande précision.

Notamment, ces indices de confiance sont calculés essentiellement à partir des probabilités de transition d'un modèle à l'autre. Ainsi, ces indices de confiance sont directement calculés pour un segment de trames acoustiques entraînant un niveau de précision peu élevé.
15

De manière classique, ces indices de confiance permettent uniquement de rejeter certaines associations qui sont corrigées manuellement par des spécialistes lors d'une étape 10 de correction longue et coûteuse.

Il apparaît donc que dans les procédés de l'état de l'art, la précision
20 des indices de confiance est insuffisante rendant ainsi les procédés de traitement longs et coûteux de par la nécessité d'interventions humaines de correction.

Le but de la présente invention est de remédier à ce problème, en définissant un procédé automatique de traitement comportant un indice de confiance de précision accrue.

25 L'invention a pour objet un procédé de traitement automatique d'un signal de parole comportant :

- une étape automatique de détermination d'au moins une séquence de modèles probabilistes issus d'un répertoire fini de modèles, chaque séquence décrivant la probabilité de réalisation acoustique d'une séquence d'unités
30 symboliques de nature phonologique issues d'un alphabet fini, ladite séquence d'unités symboliques correspondant à au moins un texte donné et lesdits modèles probabilistes comportant chacun un processus aléatoire observable correspondant à la réalisation acoustique d'unités symboliques et un processus

aléatoire non observable possédant des propriétés probabilistes connues dites propriétés de Markov ;

- une étape de détermination d'une séquence de trames de données numériques, dites trames acoustiques, représentatives de propriétés acoustiques d'un signal de parole ;

- une étape d'alignement entre ladite séquence de trames acoustiques et ladite séquence de modèles, chaque modèle étant associé à une sous-séquence de trames acoustiques, formant un segment acoustique, et chaque valeur du processus non observable de chaque modèle étant associée à une sous-séquence de trames acoustiques formant un sous-segment acoustique pour délivrer une séquence de valeurs de processus non observables associant une valeur à chaque trame acoustique, dite séquence alignée ; et

- une étape de détermination d'un indice de confiance d'alignement acoustique pour chaque association entre un modèle et un segment acoustique, dit indice de confiance d'alignement de modèle et correspondant à une estimation de la probabilité a posteriori du modèle étant donné l'observation du segment acoustique correspondant, dite probabilité de modèle a posteriori.

Chaque étape de détermination d'un indice de confiance d'alignement pour un modèle comprend le calcul de la valeur dudit indice au moins à partir d'une combinaison :

- de la probabilité d'observation de chaque trame acoustique étant donnée la valeur du processus non observable, dite probabilité de modèle et déterminée à partir de propriétés mathématiques connues du modèle et de ladite séquence de trames acoustiques ;
- de probabilités de réalisation a priori de tous les modèles dudit répertoire, indépendamment les uns des autres, dites probabilités de modèles a priori ; et
- de l'estimation analytique de la durée moyenne d'occupation des valeurs du processus non observable du modèle.

Suivant d'autres caractéristiques :

- chaque étape de détermination d'un indice de confiance acoustique pour un modèle comporte une sous-étape de détermination de l'estimation de la probabilité a priori de chaque valeur du processus non observable du modèle, dite probabilité de valeur a priori, réalisée à partir de ladite estimation analytique

de la durée moyenne d'occupation des valeurs du processus non observable du modèle ;

- chaque étape de détermination d'un indice de confiance d'alignement pour un modèle comporte une sous-étape de détermination d'un indice de confiance pour chaque trame acoustique formant le segment acoustique associé
5 audit modèle et une sous-étape de combinaison des indices de confiance de chaque trame dudit segment afin de délivrer ledit indice de confiance dudit modèle ;

- chaque sous-étape de détermination d'un indice de confiance pour
10 une trame donnée comporte :

- une sous-étape de calcul initial combinant la probabilité de modèle, la probabilité de modèle a priori du modèle en cours et la durée moyenne d'occupation des valeurs non observables pour toutes les valeurs du processus non observable de ladite séquence alignée et du modèle en cours ;

- une sous-étape de calcul du produit de la probabilité de modèle, de la probabilité de modèle a priori et de la probabilité de valeur a priori, réalisée pour chaque valeur du processus non observable de tous les modèles possibles dudit répertoire fini de modèles ; et

- une sous-étape de sommation de tous lesdits produits
20 pour tous les modèles possibles dudit répertoire fini de modèles, afin de délivrer ledit indice de confiance de ladite trame acoustique donnée à partir des résultats desdites sous-étapes ;

- ladite étape automatique de détermination d'une séquence de modèles probabilistes correspondant à un texte donné comporte :

- une sous-étape d'acquisition d'une représentation graphémique dudit texte donné ;

- une sous-étape de détermination d'une séquence d'unités symboliques issues d'un alphabet symbolique fini à partir de ladite représentation graphémique ; et

- une sous-étape automatique de modélisation de ladite sé-
30 quence d'unités par sa décomposition sur une base desdits modèles probabilistes afin de délivrer ladite séquence de modèles probabilistes ;

- ladite sous-étape de modélisation associe un unique modèle probabi-
liste à chaque unité symbolique de ladite séquence d'unités symboliques ;

- ladite étape de détermination d'une séquence de trames numériques comporte :

5 - une sous-étape d'acquisition d'un signal de parole correspondant à l'élocution dudit texte donné, adaptée pour délivrer une séquence d'échantillons numériques dudit signal de parole ; et

 - une sous-étape d'analyse spectrale desdits échantillons pour délivrer une décomposition du spectre fréquentiel dudit signal de parole sur une échelle non linéaire, ladite décomposition formant ladite séquence de trames acoustiques ;

10 - ladite sous-étape d'analyse spectrale correspond à une sous-étape de transformation de Fourier dudit signal de parole, de détermination de la distribution de son énergie sur une échelle non linéaire par filtrage, et de transformation en cosinus ;

 - ladite étape d'alignement entre ladite séquence de trames acoustiques et ladite séquence de modèles comporte :

15 - une sous-étape de calcul d'une pluralité d'alignements possibles chacun associé à un indice de pertinence ; et

 - une sous-étape de sélection d'un unique alignement parmi ladite pluralité d'alignements possibles ;

20 - ladite sous-étape de détermination d'une pluralité d'alignements possibles comprend le calcul d'au moins un alignement optimal, tel que déterminé par un algorithme dit de Viterbi ;

 - il comporte en outre une étape de modification locale de ladite séquence de modèles, en fonction desdits indices de confiance d'alignement déterminés pour chaque modèle de ladite séquence de modèles ;

25 - ladite étape de modification locale comprend une sous-étape de suppression d'un modèle de ladite séquence de modèles ;

 - ladite étape de modification locale comporte une sous-étape de substitution d'un modèle de ladite séquence de modèles par un autre modèle ;

30 - ladite étape de modification locale comporte une sous-étape d'insertion d'un modèle entre deux modèles de ladite séquence de modèles ;

 - lesdites étapes d'alignement et de calcul d'un indice de confiance sont répétées après chaque étape de modification locale de ladite séquence de modèles ;

- ladite étape de détermination d'au moins une séquence de modèles est adaptée pour la détermination d'une séquence de modèles correspondant à un texte donné et en ce que ladite séquence de trames acoustiques est représentative de propriétés d'un signal de parole correspondant à la locution
5 dudit même texte donné.

- ladite étape de détermination de séquences de modèles est adaptée pour la détermination d'une pluralité de séquences de modèles correspondant chacun à un texte donné et en ce que ladite séquence de trames acoustiques est représentative de propriétés d'un signal de parole correspondant à la locution
10 d'un texte quelconque, ledit procédé comportant une étape de sélection d'une ou plusieurs séquences de modèles parmi ladite pluralité, pour la réalisation de ladite étape de détermination d'indices de confiance.

- lesdits modèles sont des modèles dont les processus observables sont à valeurs discrètes, les valeurs des processus non observables étant les
15 états de ces processus ;

- lesdits modèles sont des modèles dont les processus non observables sont à valeurs continues.

L'invention sera mieux comprise à la lecture de la description qui va suivre, donnée uniquement à titre d'exemple et faite en se référant aux dessins annexés, sur lesquels, outre la Fig.1 déjà citée représentant un organigramme
20 d'un procédé de traitement automatique de l'état de l'art dans le cadre de la synthèse vocale :

- la Fig.2 représente un organigramme d'un procédé de traitement selon l'invention dans le cadre de la synthèse vocale ; et

25 - la Fig.3 représente le détail de signaux spécifiques au cours du procédé décrit en référence à la figure 2.

Le procédé de l'invention décrit sur la figure 2, comporte une étape 20 de détermination automatique d'une séquence de modèles probabilistes représentative d'un texte donné.

30 Dans le mode de réalisation décrit, cette étape 20 comporte une sous-étape 22 d'acquisition d'une représentation symbolique d'un texte donné, telle qu'une représentation graphémique ou orthographique.

Par exemple, cette représentation graphémique est un texte rédigé à l'aide de l'alphabet latin, désigné par la référence TXT sur la figure 3.

Le procédé comporte ensuite une sous-étape 24 de détermination d'une séquence d'unités symboliques de nature phonologique d'un alphabet fini, à partir de ladite représentation graphémique.

Une telle séquence d'unités symboliques, repérée par la référence U
5 sur la figure 3, est par exemple composée de phonèmes extraits d'un alphabet phonétique.

Cette sous-étape 24 est réalisée automatiquement aux moyens de techniques classiques de l'état de l'art, telles que phonétisation ou autres.

Ainsi, par exemple, le texte « monsieur » en langue française est re-
10 présenté à l'issue de la sous-étape 24, par la séquence d'unités phonétiques : [m]-[ø]-[s]-[j]-[ø].

Notamment, cette sous-étape 24 met en œuvre un système de phoné-
tisation automatique utilisant des bases de données et permettant de décompo-
ser n'importe quel texte sur un alphabet symbolique fini.

15 L'étape 20 comporte ensuite une sous-étape 26 automatique de mo-
délisation de la séquence U d'unités phonétiques par sa décomposition sur une
base de modèles probabilistes de modèles de Markov cachés, couramment dési-
gnés HMM.

Dans le mode de réalisation décrit, les modèles de la séquence sont
20 référencés H_1 à H_N sur la figure 3 et sont des modèles discrets comportant cha-
cun un processus aléatoire observable correspondant à une réalisation acousti-
que et un processus aléatoire non observable désigné Q et possédant des pro-
priétés probabilistes connues dites propriétés de Markov, selon lesquelles la ré-
alisation de l'état futur d'un processus aléatoire ne dépend que de l'état présent
25 de ce processus.

Ces modèles sont définis préalablement, par exemple à l'aide de ré-
seaux neuronaux permettant de déterminer l'ensemble de leurs paramètres et
notamment, des probabilités de maintien dans un état caché donné ainsi que des
probabilités de transition entre l'état caché final d'un modèle et l'état caché initial
30 d'un autre modèle.

Les propriétés mathématiques connues des modèles de Markov per-
mettent de déterminer la probabilité d'observation d'une réalisation acoustique
désignée O_i , étant donné l'état du processus non observable Q, dite probabilité
de modèle, notée P_m , et correspondant à :

$$P_m = P(O_t | Q_t)$$

On rappelle qu'une telle expression est une probabilité conditionnelle et correspond à la probabilité d'observation de la variable aléatoire O_t étant supposé réalisé un état donné Q_t du processus aléatoire Q .

5 Ces modèles de Markov sont issus d'un répertoire fini comportant par exemple 36 modèles différents référencés λ_1 à λ_I et sont associés aux unités symboliques de l'alphabet symbolique cité précédemment.

Dans le mode de réalisation décrit, chaque unité symbolique est associée à un unique modèle de Markov caché, de sorte que la séquence U d'unités phonétiques permet de déterminer directement une séquence H_1 à H_N , notée
10 H_1^N , de modèles de Markov cachés décrivant la probabilité de réalisation acoustique de la séquence U d'unités symboliques.

Ainsi, l'étape 20 permet la détermination automatique d'une séquence H_1^N de modèles correspondant à la phonétisation automatique d'un texte donné.

15 Parallèlement à l'étape 20 de détermination de la séquence H_1^N de modèles, le procédé comporte de manière classique, une étape 40 de détermination d'une séquence de trames numériques, dites trames acoustiques, représentative de propriétés acoustiques d'un signal de parole correspondant à l'élocution du texte donné TXT.

20 Dans le mode de réalisation décrit, cette étape 40 comporte une sous-étape 42 d'acquisition d'un signal de parole, identifié par la référence $s(t)$ sur la figure 3, et correspondant à l'élocution du texte donné TXT.

Cette sous-étape 42 permet l'acquisition de la forme temporelle du signal de parole $s(t)$ lequel est numérisé et échantillonné de sorte que la sous-
25 étape 42 délivre une séquence d'échantillons numériques du signal de parole $s(t)$.

Ainsi que cela a été dit précédemment, le signal de parole $s(t)$ est directement lié aux caractéristiques d'élocution du locuteur de sorte que des variations significatives peuvent apparaître entre différentes locutions et qu'une pluralité de signaux acoustiques peuvent être considérés comme représentatifs du
30 même texte TXT.

L'étape 40 comporte ensuite une sous-étape 44 d'analyse spectrale des échantillons numériques du signal de parole $s(t)$ pour en délivrer une décomposition du spectre fréquentiel.

5 De manière classique, cette analyse spectrale est une analyse dite «MFCC» (Mel Frequency Cepstrum Coefficient) qui tient compte des propriétés non linéaires de la perception auditive et d'une déconvolution entre l'onde acoustique et les caractéristiques de timbre.

10 Dans le mode de réalisation décrit, cette analyse est réalisée sur une fenêtre glissante de type Hamming dont le résultat forme une séquence, référencée O_1^T sur la figure 3, de trames acoustiques ou vecteurs acoustiques référencés O_1 à O_T .

15 La sous-étape 44 d'analyse spectrale correspond par exemple à une transformation de Fourier du signal de parole $s(t)$, à une détermination de la distribution de son énergie sur une échelle non linéaire par filtrage puis à une transformation en cosinus.

Le procédé comporte ensuite une étape 60 d'alignement entre la séquence O_1^T de trames acoustiques et la séquence H_1^N de modèles probabilistes.

Notamment, cette étape 60 d'alignement permet la sélection d'un alignement optimal au sens de l'algorithme dit de Viterbi.

20 Ainsi, cette étape 60 d'alignement comporte une sous-étape 62 de calcul d'une pluralité d'alignements possibles, chacun associé à un indice de vraisemblance et une sous-étape 64 de sélection d'un unique alignement parmi ladite pluralité d'alignements possibles.

25 De telles techniques d'alignements sont connues de l'état de l'art et permettent de délivrer une séquence de trames acoustiques étiquetées de sorte que chaque modèle H_n de la séquence de modèles H_1^N est associé à une sous-séquence $O(H_n)$ de trames acoustiques formant un segment acoustique.

30 De même, chaque état du processus non observable Q de chaque modèle H_n est associé à une sous-séquence de trames acoustiques formant un sous-segment acoustique ainsi que cela est représenté en référence à la figure 3.

Ainsi, une étiquette de début et une étiquette de fin sont déterminées pour chaque segment acoustique $O(H_n)$ de la séquence O_1^T , associé à un modèle H_n donné de la séquence H_1^N .

5 Cette étape 60 permet également de délivrer une séquence \tilde{Q}_t d'états non observable, dite séquence alignée, associant à chaque trame acoustique O_t un état non observable donné d'un modèle donné, noté q_j^i et correspondant au jème état de ième modèle de la séquence ainsi que cela est représenté sur la figure 3.

10 Le procédé comporte ensuite une étape 80 de détermination d'un indice de confiance d'alignement acoustique pour chaque association entre un modèle H_n et un segment acoustique $O(H_n)$.

Cet indice de confiance est appelé indice de confiance d'alignement de modèle, noté I_n , et correspond à une estimation de la probabilité a posteriori du modèle étant donné l'observation du segment acoustique correspondant noté P_{mp} et correspondant à :

$$P_{mp} = P(H_n | O(H_n))$$

Dans le cadre de l'invention, chaque étape 80 de détermination d'un indice de confiance d'alignement I_n pour un modèle H_n est réalisé à partir d'une combinaison :

20 - de la probabilité d'observation de chaque trame acoustique étant donné la valeur du processus non observable du modèle correspondant, soit encore la probabilité de modèle P_m définie précédemment ;

- de probabilités de réalisation a priori de tous les modèles λ_1 à λ_I du répertoire, indépendamment les uns des autres, dites probabilités de modèle a priori et notées $P(\lambda_i)$; et

25 - du temps moyen de séjour sur chacun des états q_j^i du modèle H_n , noté $\bar{d}(q_j^i)$, calculé à partir de paramètres caractéristiques du modèle H_n et en particulier des paramètres de transition entre les états non observables.

La probabilité de modèles P_m est déterminée à partir des propriétés probabilistes connues du modèle H_n et de la séquence de trames acoustiques observée O_1^T .

Les probabilités de modèle a priori $P(\lambda_i)$ sont, par exemple, estimées
5 préalablement par comptage des occurrences de phonèmes à partir de transcriptions graphémiques et/ou phonétiques.

Le temps moyen de séjour permet notamment l'estimation de la probabilité a priori de chaque valeur ou état du processus non observable Q d'un modèle H_n , dite probabilité de valeur a priori notée P_{vp} qui s'exprime sous la
10 forme de probabilités conditionnelles par :

$$P_{vp} = P(q_j^i | \lambda_i)$$

et qui correspond à la probabilité a priori d'être dans un état non observable donné référencé q_j^i d'un modèle λ_i donné, noté q_j^i ainsi que cela a été décrit précédemment.

15 Les séquences étant dans des relations d'ordre temporel, les probabilités P_{vp} peuvent être exprimées de manière analytique par le ratio entre le temps moyen passé sur un état q_j^i , noté $\bar{d}(q_j^i)$, et le temps moyen d'occupation du modèle λ_i noté $\bar{d}(\lambda_i)$ et correspondant à la sommes des temps moyens de séjour sur chacun des états le constituant.

20 On peut alors écrire la relation analytique générale suivante :

$$P_{vp} \# \frac{\bar{d}(q_j^i)}{\bar{d}(\lambda_i)}$$

En fonction des modes de réalisation, le procédé de l'invention peut recevoir directement les P_{vp} , par exemple calculés préalablement et stockés dans une mémoire, ou encore recevoir les estimations $\bar{d}(q_j^i)$ de la durée moyenne
25 d'occupation des états des processus non observables du modèle et effectuer le calcul lors d'une sous-étape de détermination de la probabilité de valeur a priori P_{vp} .

Il apparaît alors que l'indice de confiance I_n peut s'exprimer selon la relation suivante :

30

$$I_n = \log \left[\frac{\prod_{t=b(n)}^{e(n)} \frac{P(O_t | \tilde{Q}_t) \bar{d}(\tilde{Q}_t)}{\sum_{i=1}^I \sum_{j=1}^{J(i)} P(O_t | Q_t = q_j^i) \bar{d}(q_j^i) P(\lambda_i)}}{P(H_n)} \right]$$

Dans cette relation, \tilde{Q}_t correspond à l'instant de la séquence alignée délivrée à l'issue de la sous-étape 64 et donc dans le mode de réalisation décrit, à une séquence d'états optimale au sens de l'algorithme de Viterbi s'étendant
 5 entre les instants $t=b(n)$ et $t=e(n)$ correspondant respectivement au début et à la fin de la séquence d'observation $O(H_n)$.

Le terme $\bar{d}(\tilde{Q}_t)$ correspondant à la durée moyenne de l'état à l'instant t de la séquence alignée et le terme $\bar{d}(H_n)$ correspondant à la durée moyenne du même modèle de la séquence H_1^N , ils sont tous deux obtenus à partir de la durée moyenne d'occupation des états non observables notée $\bar{d}(q_j^i)$ de manière
 10 générale.

Enfin, l'indice i permet de parcourir les modèles λ_1 à λ_I du répertoire de modèles et l'indice j permet de parcourir les états non observables 1 à $J(i)$ de chaque modèle.

15 Afin de mettre en oeuvre cette relation, l'étape 80 comporte une sous-étape 82 de calcul initial au cours de laquelle le numérateur de la relation est calculé pour une trame donnée.

Au cours de cette sous-étape 82 la probabilité de modèle $P_m = P(O_t | \tilde{Q}_t)$ est combinée avec la probabilité de modèle a priori du modèle en
 20 cours $P(H_n)$, la durée moyenne d'occupation de la séquence alignée $\bar{d}(\tilde{Q}_t)$ et la durée moyenne du modèle en cours $\bar{d}(H_n)$.

L'étape 80 comporte ensuite une sous-étape 84 de calcul du produit de la probabilité de modèles P_m avec la probabilité de modèle a priori $P(\lambda_i)$ et la probabilité de valeurs a priori P_{vp} . Cette sous-étape 84 est réalisée pour tous les
 25 états non observables de tous les modèles possibles du répertoire fini de modèles.

Par la suite, le procédé comporte une étape 86 de sommation de tous les produits déterminés précédemment pour tous les modèles possibles λ_1 à λ_I du répertoire fini de modèles.

On détermine ainsi un indice de confiance pour une trame acoustique donnée.

Le procédé comporte ensuite une étape 88 de combinaison des indices de confiance de chaque trame du segment acoustique donné afin de délivrer l'indice de confiance I_n du modèle H_n considéré.

La relation définissant I_n peut se ramener à l'équation algorithmique suivante :

$$I_n = \sum_{t=b(n)}^{e(n)} \left[\log P(O_t | \tilde{Q}_t) + \log \bar{d}(\tilde{Q}_t) - \log \bar{d}(H_n) + \log P(H_n) - \log \left[\sum_{i=1}^I \sum_{j=1}^{J(i)} P(O_t | Q_t = q_j^i) \frac{\bar{d}(q_j^i)}{\bar{d}(\lambda_i)} P(\lambda_i) \right] \right]$$

De ce fait, les sous-étapes 82, 84 et 86 de l'étape 80 peuvent être décrites de manière algorithmique sous la forme suivante :

On définit un accumulateur PO puis on effectue les calculs suivants :

PO = 0 ;

Pour chaque modèle λ_i du répertoire, avec i compris entre 1 et I et pour chaque état j du modèle λ_i , avec j compris entre 1 et J(i), on calcule :

$$PO = PO + P(O_t | q_j^i) \frac{\bar{d}(q_j^i)}{\bar{d}(\lambda_i)} P(\lambda_i) ;$$

$$I_n = I_n + \log P(O_t | \tilde{Q}_t) ;$$

$$I_n = I_n + \log \bar{d}(q_t^n) ;$$

$$I_n = I_n - \log \bar{d}(H_n) ;$$

$$I_n = I_n + \log P(H_n) ;$$

$$I_n = I_n - PO.$$

Dans les relations précédentes, la valeur de l'indice précédente I_n est successivement instanciée à la valeur courante.

Avantageusement, le procédé comporte ensuite une sous-étape 90 de normalisation de l'indice de confiance d'alignement afin de délivrer un indice de confiance normalisé par rapport à la durée totale du modèle.

5 Ainsi, l'étape 80 délivre à partir de la probabilité de modèles P_m de la probabilité de modèle a priori $P(\lambda_i)$ et de la durée moyenne d'occupation des états non observables $\bar{d}(q_j^i)$ l'indice de confiance d'alignement I_n .

10 Cet indice de confiance est d'une très grande fiabilité du fait notamment qu'il est calculé à partir de probabilités de valeur a priori P_{vp} estimée de manière analytique à partir de la durée moyenne d'occupation des états du processus non observable permettant ainsi de prendre en compte le temps passé dans chacun des états cachés et donc de prendre en compte les caractéristiques temporelles de la séquence de modèles H_1^N .

15 Par la suite, le procédé comporte une étape 100 de modification locale de la séquence H_1^N en fonction des indices de confiance d'alignement I_n déterminés pour chaque modèle de la séquence.

Ainsi, lors de l'étape 100, une décision permettant soit d'accepter soit de rejeter un modèle est prise pour chaque modèle de la séquence H_1^N .

20 Dans le mode de réalisation décrit, cette décision est prise automatiquement en fonction de la mesure de confiance I_n déterminée lors de l'étape 80 et d'un point de fonctionnement considéré comme un seuil d'acceptation ou de rejet du modèle.

Lorsque un modèle est rejeté, l'étape 100 de modification locale comporte alors alternativement une sous-étape de suppression, de substitution ou d'insertion d'un modèle.

25 Ainsi, un ou plusieurs modèles de la séquence de modèles H_1^N peuvent, manuellement ou automatiquement, être supprimés, remplacés ou encore un ou plusieurs nouveaux modèles peuvent être intercalés entre deux modèles de la séquence.

30 Avantageusement, la séquence ainsi modifiée est alors utilisée à nouveau dans le procédé pour être à nouveau alignée avec la séquence de trames acoustiques O_1^T lors de l'étape 60 et donner lieu à un nouveau calcul d'un indice

de confiance pour chaque association entre un modèle et un segment acoustique lors de l'étape 80.

Les étapes 60, 80 et 100 sont réitérées jusqu'à ce qu'il n'y ait plus de modèle rejeté ou encore qu'il n'y ait plus de modifications possibles de sorte que
5 la séquence de trames étiquetées délivrée correspond à la meilleure hypothèse de décodage possibles.

Il apparaît donc que le procédé de l'invention permet la définition d'un indice de confiance d'une précision accrue. L'utilisation de cet indice permet notamment d'automatiser l'ensemble du procédé de traitement d'un signal de parole, de définir des modifications automatiques et d'obtenir un résultat optimisé.
10

Le procédé décrit peut être mis en oeuvre par des moyens logiciels et/ou matériels tels que des ordinateurs, des micro-processeurs, ou tout autre équipement adapté.

Le procédé décrit peut par exemple être utilisé dans des systèmes de
15 synthèse vocale ou de constitution de bases de données d'apprentissage pour des systèmes de reconnaissance vocale et permet, du fait de l'utilisation d'un indice de confiance de grande précision et du rebouclage après une modification automatique, d'obtenir des séquences de trames étiquetées fiables et exploitables directement sans requérir d'intervention humaine.

20 Dans le cadre de la synthèse vocale, ainsi que cela a été décrit, la séquence de modèles et la séquence de trames acoustiques correspondent au même texte.

En variante, le procédé de l'invention peut être utilisé dans des systèmes de reconnaissance vocale par exemple pour former des bases de données
25 d'échantillons de reconnaissance ou encore pour permettre de reconnaître un énoncé dans un répertoire de séquences de modèles.

Par exemple, les séquences candidates de modèle prédéterminées sont alignées successivement avec la séquence de trames acoustiques dite séquence cible et les indices de confiance de chaque modèle sont combinés sur
30 l'ensemble de la séquence afin de délivrer une mesure de similitude entre la séquence de modèles et la séquence de trames acoustiques. L'association ayant la mesure de similitude la plus élevée est retenue.

Alternativement, la séquence sélectionnée est obtenue par modification à chaque rebouclage de manière similaire au mode de réalisation décrit précédemment.

5 Enfin, dans l'exemple décrit, les modèles de Markov cachés sont des modèles dont les processus non observables sont à états discrets. Cependant, le procédé peut également être réalisé avec des modèles dont les processus non observables sont à valeurs continues.

REVENDEICATIONS

1. Procédé de traitement automatique d'un signal de parole comportant :

- une étape (20) automatique de détermination d'au moins une séquence (H_1^N) de modèles probabilistes (H_n) issus d'un répertoire fini de modèles $(\lambda_1 \text{ à } \lambda_I)$, chaque séquence (H_1^N) décrivant la probabilité de réalisation acoustique d'une séquence (U) d'unités symboliques de nature phonologique issues d'un alphabet fini, ladite séquence (U) d'unités symboliques correspondant à au moins un texte donné (TXT) et lesdits modèles probabilistes (λ_i) comportant chacun un processus aléatoire observable correspondant à la réalisation acoustique d'unités symboliques et un processus aléatoire non observable (Q) possédant des propriétés probabilistes connues dites propriétés de Markov ;
- une étape (40) de détermination d'une séquence (O_1^T) de trames de données numériques, dites trames acoustiques, représentatives de propriétés acoustiques d'un signal de parole $(s(t))$;
- une étape (60) d'alignement entre ladite séquence (O_1^T) de trames acoustiques et ladite au moins une séquence (H_1^N) de modèles, chaque modèle (H_n) étant associé à une sous-séquence $(O(H_n))$ de trames acoustiques, formant un segment acoustique, et chaque valeur du processus non observable (Q) de chaque modèle (H_n) étant associée à une sous-séquence de trames acoustiques formant un sous-segment acoustique, pour délivrer une séquence (\tilde{Q}_t) de valeurs de processus non observables associant une valeur à chaque trame acoustique, dite séquence alignée (\tilde{Q}_t) ; et
- une étape (80) de détermination d'un indice de confiance (I_n) d'alignement acoustique pour chaque association entre un modèle (H_n) de la séquence et un segment acoustique $(O(H_n))$, dit indice (I_n) de confiance d'alignement de modèle et correspondant à une estimation de la probabilité a posteriori du modèle (H_n) étant donné l'observation du segment acoustique correspondant $(O(H_n))$, dite probabilité de modèle a posteriori (P_{mp}) ,

caractérisé en ce que chaque étape (80) de détermination d'un indice de confiance d'alignement (I_n) pour un modèle (H_n) comprend le calcul de la valeur dudit indice (I_n) au moins à partir d'une combinaison :

- de la probabilité d'observation de chaque trame acoustique étant donnée la valeur du processus non observable (Q), dite probabilité de modèle (P_m) et déterminée à partir de paramètres caractéristiques connus du modèle probabiliste (H_n) ;
- de probabilités de réalisation a priori de tous les modèles dudit répertoire indépendamment les uns des autres, dites probabilités de modèle a priori ($P(\lambda_i)$) ; et

- de l'estimation analytique ($\bar{d}(q_j^i)$) de la durée moyenne d'occupation des valeurs du processus non observable du modèle.

2. Procédé selon la revendication 1, caractérisé en ce que chaque étape (80) de détermination d'un indice de confiance acoustique (I_n) pour un modèle comporte une sous-étape de détermination de l'estimation de la probabilité a priori de chaque valeur du processus non observable (Q) du modèle (H_n), dite probabilité de valeur a priori (P_{vp}), réalisée à partir de ladite estimation analytique ($\bar{d}(q_j^i)$) de la durée moyenne d'occupation des valeurs du processus non observable du modèle.

3. Procédé selon l'une quelconque des revendications 1 ou 2, caractérisé en ce que chaque étape (80) de détermination d'un indice de confiance d'alignement (I_n) pour un modèle (H_n) comporte une sous-étape de détermination d'un indice de confiance pour chaque trame acoustique formant le segment acoustique associé audit modèle (H_n) et une sous-étape (88) de combinaison des indices de confiance de chaque trame dudit segment ($O(H_n)$) afin de délivrer ledit indice de confiance (I_n) dudit modèle (H_n).

4. Procédé selon la revendication 3, caractérisé en ce que chaque sous-étape de détermination d'un indice de confiance pour une trame donnée comporte :

- une sous-étape (82) de calcul initial combinant la probabilité de modèle (P_m), la probabilité de modèle a priori du modèle en cours ($P(H_n)$) et la durée moyenne d'occupation des valeurs non observables pour toutes les

valeurs du processus non observable de ladite séquence alignée (\tilde{Q}_l) et du modèle en cours (H_n) ;

- une sous-étape (84) de calcul du produit de la probabilité de modèle (P_m), de la probabilité de modèle a priori et de la probabilité de valeur a priori (P_{vp}), réalisée pour chaque valeur du processus non observable (Q) de tous les modèles possibles (λ_1 à λ_I) dudit répertoire fini de modèles ; et

- une sous-étape (86) de sommation de tous lesdits produits pour tous les modèles (λ_1 à λ_I) possibles dudit répertoire fini de modèles, afin de délivrer ledit indice de confiance de ladite trame acoustique donnée à partir des résultats desdites sous-étapes (82, 84, 86).

5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce que il comporte une sous-étape (90) de normalisation des indices de confiance par modèle (ln) en fonction de la durée des modèles.

- 6. Procédé selon l'une quelconque des revendications 1 à 5, caractérisé en ce que ladite étape (20) automatique de détermination d'une séquence de modèles probabilistes correspondant à un texte donné (TXT) comporte :

- une sous-étape (22) d'acquisition d'une représentation graphémique dudit texte donné (TXT) ;
- une sous-étape (24) de détermination de ladite séquence (U) d'unités symboliques à partir de ladite représentation graphémique ; et
- une sous-étape (26) automatique de modélisation de ladite séquence (U) d'unités symboliques par sa décomposition sur une base desdits modèles probabilistes (λ_1 à λ_I) afin de délivrer ladite séquence (H_1^N) de modèles probabilistes.

7. Procédé selon la revendication 6, caractérisé en ce que ladite sous-étape (24) de modélisation associe un unique modèle probabiliste (H_n) à chaque unité symbolique de ladite séquence (U) d'unités symboliques.

- 8. Procédé selon l'une quelconque des revendications 1 à 7, caractérisé en ce que ladite étape (40) de détermination d'une séquence de trames numériques (O_1^T) comporte :

- une sous-étape (42) d'acquisition d'un signal de parole ($s(t)$) correspondant à l'élocution dudit texte donné (TXT), adaptée pour délivrer une séquence d'échantillons numériques dudit signal de parole ($s(t)$) ; et

5 - une sous-étape (44) d'analyse spectrale desdits échantillons pour délivrer une décomposition du spectre fréquentiel dudit signal de parole ($s(t)$) sur une échelle non linéaire, ladite décomposition formant ladite séquence (O_1^T) de trames acoustiques.

9. Procédé selon la revendication 8, caractérisé en ce que ladite sous-étape (44) d'analyse spectrale correspond à une sous-étape de transformation de
10 Fourier dudit signal de parole ($s(t)$), de détermination de la distribution de son énergie sur une échelle non linéaire par filtrage, et de transformation en cosinus.

10. Procédé selon l'une quelconque des revendications 1 à 9, caractérisé en ce que ladite étape (60) d'alignement entre ladite séquence (O_1^T) de trames acoustiques et ladite séquence (H_1^N) de modèles comporte :

15 - une sous-étape (62) de calcul d'une pluralité d'alignements possibles chacun associé à un indice de pertinence ; et

 - une sous-étape (64) de sélection d'un unique alignement parmi ladite pluralité d'alignements possibles.

11. Procédé selon la revendication 10, caractérisé en ce que ladite
20 sous-étape (62) de détermination d'une pluralité d'alignements possibles comprend le calcul d'au moins un alignement optimal, tel que déterminé par un algorithme dit de Viterbi.

12. Procédé selon l'une quelconque des revendications 1 à 11, caractérisé en ce qu'il comporte en outre une étape (100) de modification locale
25 de ladite séquence de modèles, en fonction desdits indices de confiance d'alignement (I_n) déterminés pour chaque modèle (H_n) de ladite séquence de modèles (H_1^N).

13. Procédé selon la revendication 12, caractérisé en ce que ladite
30 étape (100) de modification locale comprend une sous-étape de suppression d'un modèle de ladite séquence de modèles.

14. Procédé selon la revendication 12, caractérisé en ce que ladite étape (100) de modification locale comporte une sous-étape de substitution d'un modèle de ladite séquence (H_1^N) de modèles par un autre modèle.

5 15. Procédé selon la revendication 12, caractérisé en ce que ladite étape (100) de modification locale comporte une sous-étape d'insertion d'un modèle entre deux modèles de ladite séquence (H_1^N) de modèles.

10 16. Procédé selon l'une quelconque des revendications 12 à 15, caractérisé en ce que lesdites étapes (60) d'alignement et de (80) calcul d'un indice de confiance sont répétées après chaque étape (100) de modification locale de ladite séquence de modèles.

15 17. Procédé selon l'une quelconque des revendications 1 à 16, caractérisé en ce que ladite étape (20) de détermination d'au moins une séquence de modèles est adaptée pour la détermination d'une séquence de modèles (H_1^N) correspondant à un texte donné et en ce que ladite séquence de trames acoustiques (O_1^T) est représentative de propriétés d'un signal de parole ($s(t)$) correspondant à la locution dudit même texte donné (TXT).

20 18. Procédé selon l'une quelconque des revendications 1 à 16, caractérisé en ce que ladite étape (20) de détermination de séquences de modèles est adaptée pour la détermination d'une pluralité de séquences de modèles correspondant chacun à un texte donné et en ce que ladite séquence de trames acoustiques est représentative de propriétés d'un signal de parole correspondant à la locution d'un texte quelconque, ledit procédé comportant une étape de sélection d'une ou plusieurs séquences de modèles parmi ladite pluralité, pour la réalisation de ladite étape de détermination d'indices de
25 confiance.

19. Procédé selon l'une quelconque des revendications 1 à 18, caractérisé en ce que lesdits modèles (H_n) sont des modèles dont les processus observables sont à valeurs discrètes, les valeurs des processus non observables étant les états (Q_t) de ces processus.

30 20. Procédé selon l'une quelconque des revendications 1 à 18, caractérisé en ce que lesdits modèles (H_n) sont des modèles dont les processus non observables sont à valeurs continues.

1/3

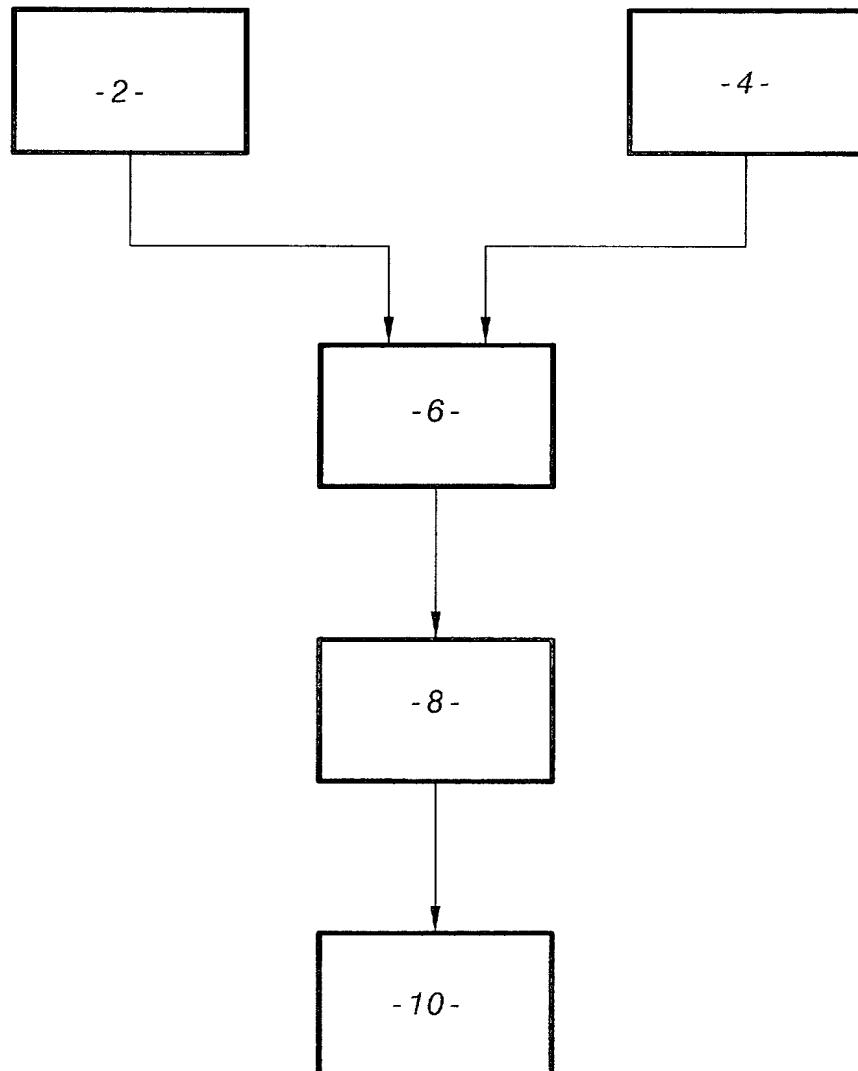


FIG.1

2/3

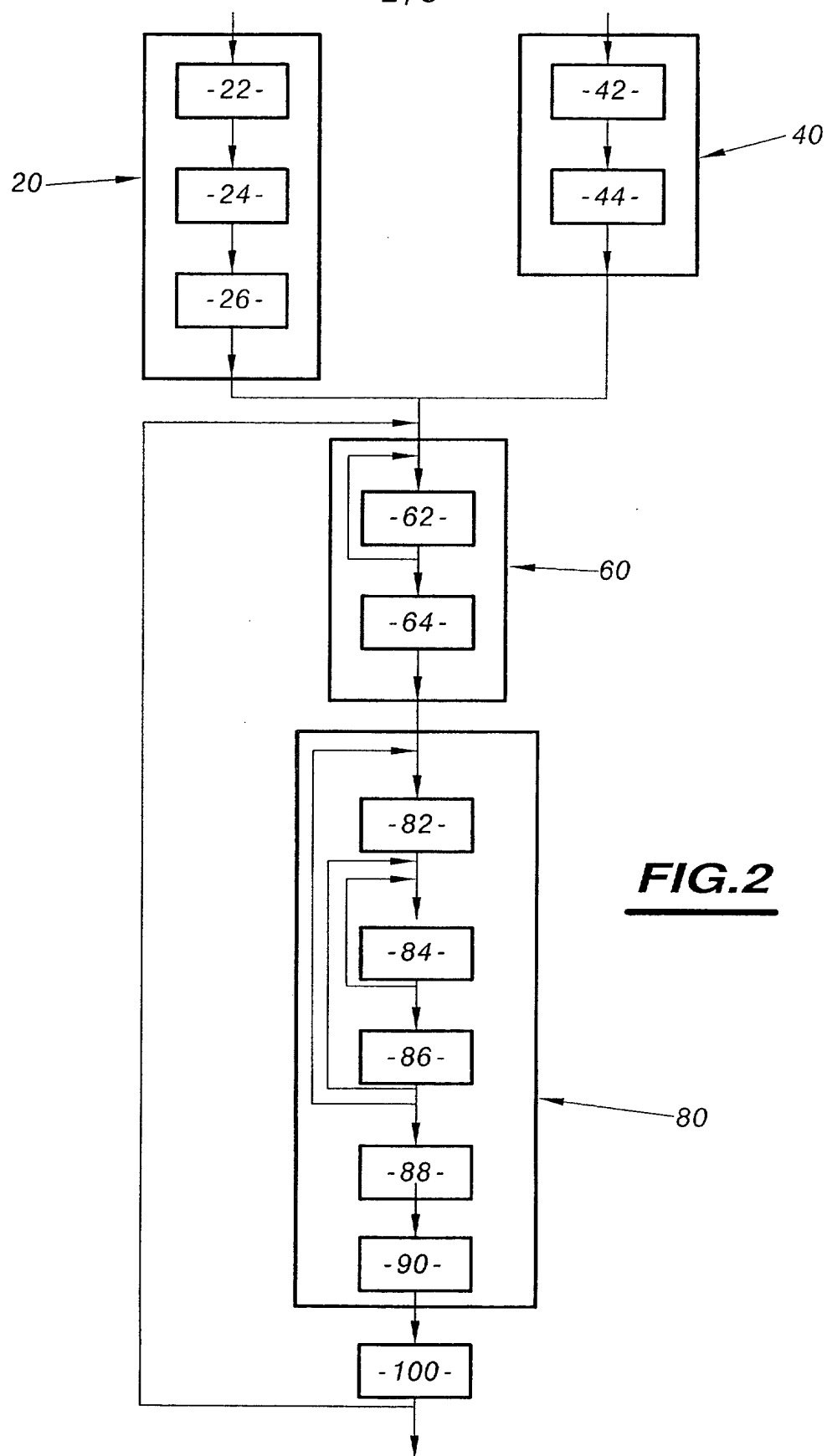


FIG.2

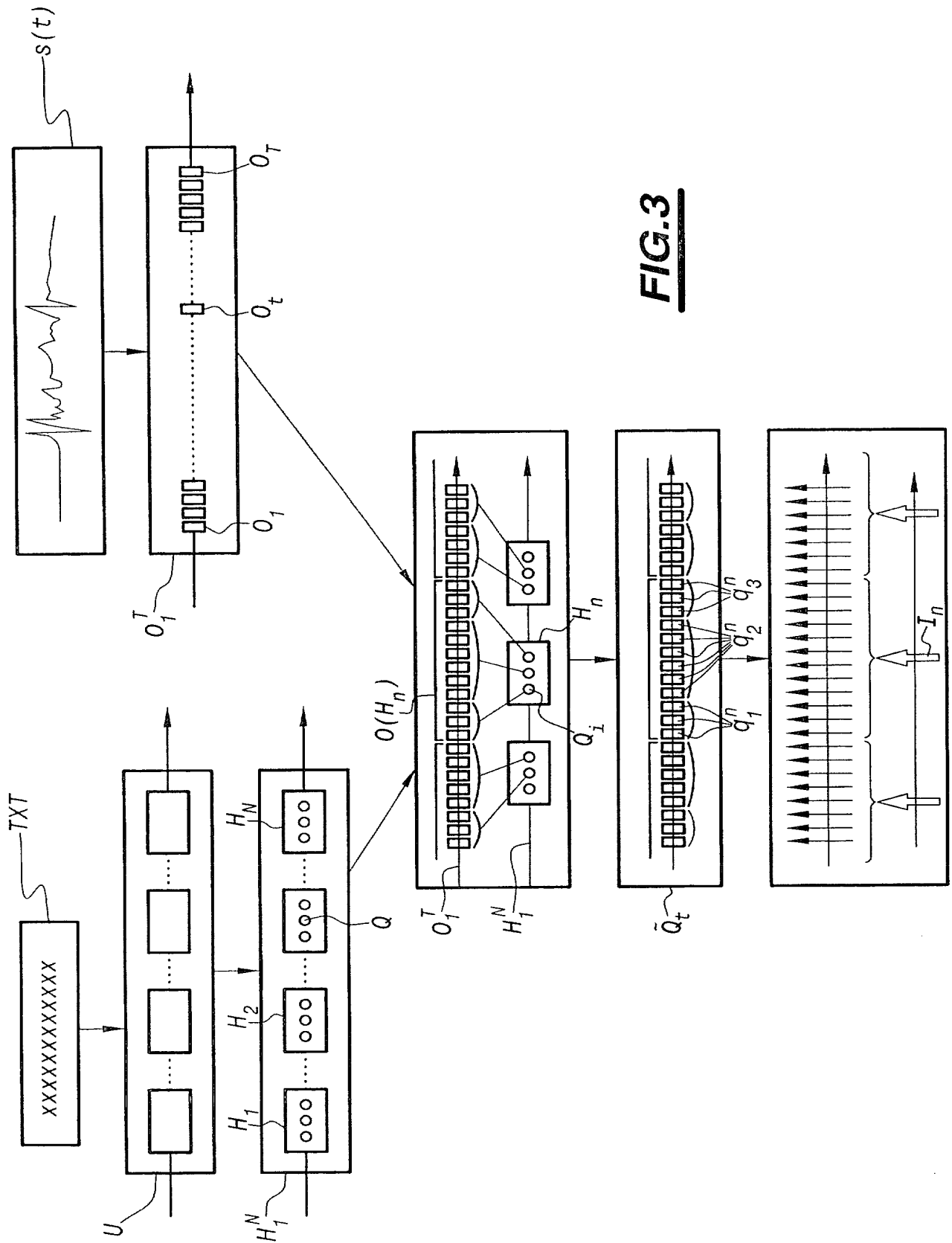


FIG. 3



DOCUMENT FILED BY:
YOUNG & THOMPSON
745 SOUTH 23RD STREET
ARLINGTON, VIRGINIA 22202
Telephone 703/521-2297